iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: IX Month of publication: September 2020

DOI: https://doi.org/10.22214/ijraset.2020.31485

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

Real Time Facial Expression Recognition using Convolution Neural Network Algorithm

Ayan Bag
Undergraduate Student, Kalyani Government Engineering College

Abstract: The most expressive way human beings display emotions is through facial expressions. The task of detecting facial expression of a human being via a computer is a very complex process due to its variability present across human faces including color, expression, position, and orientation. The aim of this paper is to presents a Convolution Neural Network (CNN) architecture for real-time facial expression recognition. We have used ICML 2013 Facial Expression Recognition Challenge dataset for this study and then trained our neural network for emotion state classification. In this study, we achieved accuracy of 84.18% and validation accuracy of 67.56% for classification of seven different emotions through facial expressions.
Keywords: Emotion Detection, Convolutional Neural Network, Image Processing, Face Detections

I. INTRODUCTION

Human communication has two main aspects: firstly Verbal Communication (auditory) and secondly, Non-Verbal Communication (Visual). The facial expression, body movements, and physiological reaction are the basic units of non-verbal communication. While communicating only a 7% effect of message is contributes by the verbal part as a whole, 38% by vocal part and 55% effect of the speaker's message is contributed by facial expression. For that reason automated and real-time facial expression would play an important role in human and human interaction. Analysis of facial expression plays fundamental roles for applications that are based on emotion recognition like Human-Computer Interaction (HCI), Social Robot, Alert System, and Pain Monitoring for the patient.

In recent years, with the popularization in deep learning, there has been great progress in the field of image classification. Convolution Neural Network (CNN) is an artificial neural network type proposed by Yann André LeCun in 1988. The Convolution Neural Network is one of the most popular deep learning architecture for image classification, recognition, and segmentation.

Convolution Neural Network is just like a human brain, made of artificial neurons and consist of hierarchical multiple hidden layers. These artificial neurons take input from an image, multiply weight, add bias, and then apply activation function. The hidden layers of a CNN typically consist of a series of convolutional layers that convolve with multiplication or other dot product. The activation function is a RELU layer and is subsequently followed by additional convolutions such as pooling layers, fully connected layers, and normalization layers referred to as hidden layers because their inputs and outputs are masked by the activation function and the final convolution. Therefore these artificial neurons can be used in image classification, image recognition, and segmentation. A better and highly accurate deep learning model can be achieved by feeding more data to the Convolution Neural Network.

Deep Learning-based facial expression recognition is one of these methods to detect human being emotional states e.g. anger, fear, happiness, sadness, surprise, and disgust. This method aims to detect facial expressions automatically to identify the human's emotional state with high accuracy. In this method, labeled facial images of various facial expressions are sent to CNN, and CNN is trained with this data-set. Then, the proposed CNN model decides which facial expression is performed.

In this study, we have proposed a CNN architecture to detect human being's facial expressions. Our proposed architecture was trained with the final dataset for the classification of 7 emotion states (happy, sad, surprised, angry, disgust, afraid, and neutral). The study aims to obtain a deep learning model that achieves a higher accuracy rate for emotion recognition through facial expression.

II. CATEGORIZING FACIAL EXPRESSION AND IT'S FEATURES

The key mechanism for describing human emotions is facial expressions. From beginning to end of the day, human changes plenty of emotions due to changes in mental and physical circumstances. Although humans are filled with various emotions, it is still can be classified into seven basic facial expression: Happiness, Sadness, Surprise, Fear, Disgust, Anger and Neutral as universal emotions. Facial Muscles movement also helps to identify human emotions. The overview of various human emotions is given in Table 1.

Table I
Universal Emotion Identification

Sl. No.	Universal Emotion Identification		
	Emotion	Definition	Motion of Facial Part
1	Anger	Anger is one of the most dangerous emotions. This emotion may be harmful so, humans are trying to avoid this emotion. Secondary emotions of anger are irritation, annoyance, frustration, hate and dislike	Eyebrows pulled down, Open eye, teeth shut and lips tightened, upper and lower lids pulled up.
2	Fear	Fear is the emotion of danger. It may be because of danger of physical or psychological harm. Secondary emotions of fear are Horror, nervousness, panic, worry and dread	Outer eyebrow down, inner eyebrow up, mouth open, jaw dropped
3	Happines s	Happiness is most desired expression by human. Secondary emotions are cheerfulness, pride, relief, hope, pleasure, and thrill.	Open Eyes, mouth edge up, open mouth, lip corner pulled up, cheeks raised, and wrinkles around eyes.
4	Sadness	Sadness is opposite emotion of Happiness. Secondary emotions are suffering, hurt, despair, pitty and hopelessness.	Outer eyebrow down, inner corner of eyebrows raised, mouth edge down, closed eye, lip corner pulled down.
5	Surprise	This emotion comes when unexpected things happens. Secondary emotions of surprise are amazement, astonishment.	Eyebrows up, open eye, mouth open, jaw dropped
6	Disgust	Disgust is a feeling of dislike. Human may feel disgust from any taste, smell, sound or tough.	Lip corner depressor, nose wrinkle ,lower lip depressor, Eyebrows pulled down

III.PROPOSED METHODOLOGY

A. Facial Expression Dataset

In this paper, I have used the ICML 2013 Facial Expression Recognition Challenge dataset. This dataset consists of 48x48 pixel gray-scale images of faces.In our dataset, 80% of the dataset was selected for training purpose and the remaining 20% was chosen for testing. Our training set contains 28,708 images and our test set contains 7178 images of 7 facial expressions (happy, sad, surprised, angry, disgust, afraid, and neutral). This dataset was prepared by Pierre-Luc Carrier and Aaron Courville, as part of their ongoing research project.

Figure II
Example of images from ICML 2013 facial expression dataset

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

B. Convolutional Neural Network Architecture

With our proposed CNN architecture, it is aimed to train the pixel value in the rectangular boxes containing facial expression quickly and to make queries with the deep learning neural network model formed. Our neural network mimics LeNet architecture and is used in the classification of 2D facial expression data. It compromised of four convolutional layers, four max-pooling layers, and two connected layers. The convolutional layers with the kernel size of 3x3 are stacked together which are followed by max-pooling layer with kernel size of 2x2. Like the convolutional layer, the RELU activation function is applied in the hidden layers of our network. RELU is applied to introduce the non-linearity of a model. After each hidden layer, a dropout layer has been inserted and the value has been set to 0.25. It means, it randomly deactivates 25% of nodes from the hidden layer to avoid over-fitting. At last, the output layer of our model consists of seven nodes as it has seven classes. *Softmax* has been used as an activation function of the output layer. *Adam* is used as our model optimizer with a learning rate of 0.0005. *Adam* is a stochastic gradient descent method that computes the individual adaptive learning rates for different parameters from estimates of first- and second-order moments of the gradients. And as loss function, *Categorical Crossentropy* is used.

The overview of the convolutional neural network architecture that has been designed for this research is given in Figure II.

Figure IIII
CNN architecture Summary

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 64)	640
batch_normalization (BatchNo	(None, 48, 48, 64)	256
activation (Activation)	(None, 48, 48, 64)	0
max_pooling2d (MaxPooling2D)	(None, 24, 24, 64)	0
dropout (Dropout)	(None, 24, 24, 64)	0
conv2d_1 (Conv2D)	(None, 24, 24, 128)	204928
batch_normalization_1 (Batch	(None, 24, 24, 128)	512
activation_1 (Activation)	(None, 24, 24, 128)	0
max_pooling2d_1 (MaxPooling2	(None, 12, 12, 128)	0
dropout_1 (Dropout)	(None, 12, 12, 128)	0
conv2d_2 (Conv2D)	(None, 12, 12, 512)	590336
batch_normalization_2 (Batch	(None, 12, 12, 512)	2048
activation_2 (Activation)	(None, 12, 12, 512)	0
max_pooling2d_2 (MaxPooling2	(None, 6, 6, 512)	0
dropout_2 (Dropout)	(None, 6, 6, 512)	0
conv2d_3 (Conv2D)	(None, 6, 6, 512)	2359808
batch_normalization_3 (Batch	(None, 6, 6, 512)	2048
activation_3 (Activation)	(None, 6, 6, 512)	0
max_pooling2d_3 (MaxPooling2	(None, 3, 3, 512)	0
dropout_3 (Dropout)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 256)	1179904
batch_normalization_4 (Batch	(None, 256)	1024
activation_4 (Activation)	(None, 256)	0
dropout_4 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 512)	131584
batch_normalization_5 (Batch	(None, 512)	2048
activation_5 (Activation)	(None, 512)	0
dropout_5 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 7)	3591

Total params: 4,478,727
Trainable params: 4,474,759
Non-trainable params: 3,968

C. Neural Network Training

In the training of our neural network, the batch size is set to 64 and the epoch number is found as 500 to converge the parameters of the neural network. The learning rate is defined as 0.0005. Our proposed CNN model consists of 4 convolutional layers with 64, 128, 512, 512 filters and kernel size are 3x3. The kernel size of our max-pooling layer is 2x2.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

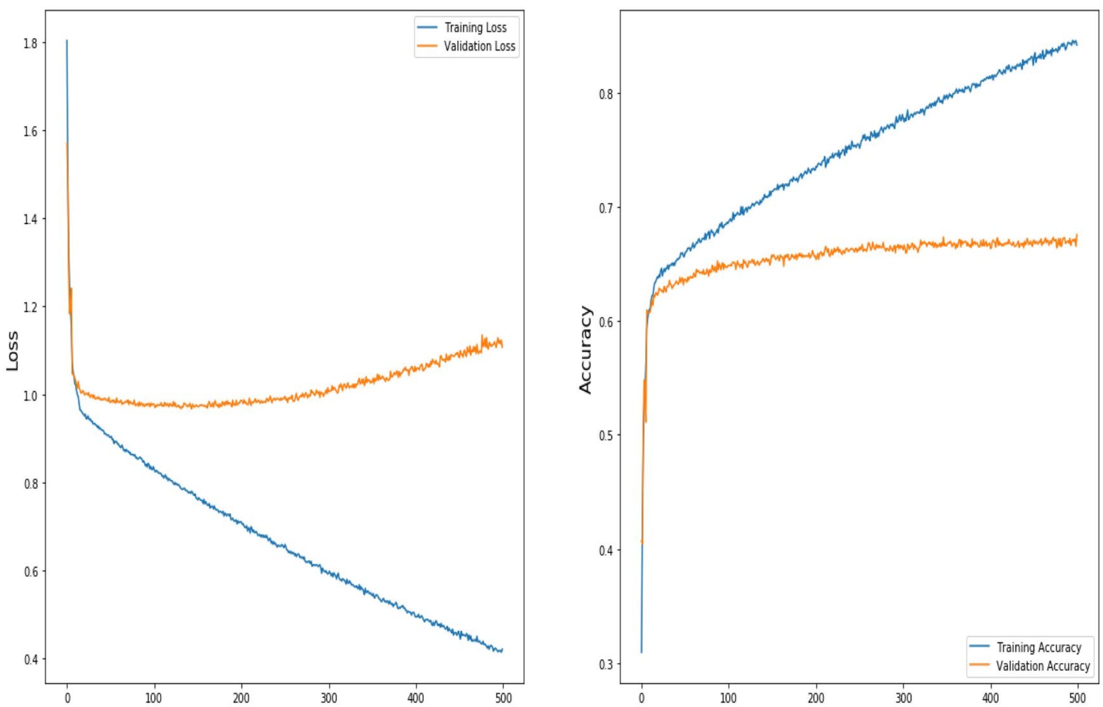
D. Real time Testing

After training of our proposed CNN architecture, the trained model is tested with real-time data. As soon as real-time images are fed to the system, it prepossesses those images. It means that, whenever an image of arbitrary size is fed to the system, it converts it 48*48 sized image. With the help of *Haar Cascade Classifier*, the model detects the faces from the image. It is mainly the region of interest which is been cropped afterward. As the model has been trained on grayscale images, the system converts the RGB image that contains 3 channels red, green and blue to gray image which consists of only 1 channel. Then to ease the classification task, the system will apply image normalization on the image. Then it will be sent to the our proposed Convolutional Neural Network architecture for classification.

IV. RESULT AND DISCUSSION

In this study, Keras and Tensorflow libraries are used to train our proposed CNN architecture and predicting the emotional states with our deep learning model. Our proposed model, has been successful to achieve training accuracy of 84.18% validation accuracy of 67.56%. Fig. 4. shows performance metrics (training accuracy and training loss, validation accuracy and validation loss) of our proposed architecture during training and testing.

Figure IIVI
Performance metrics of proposed architecture



V. CONCLUSIONS

This paper proposed an effective method for the real-time classification of seven different emotions by facial expression. In this study, facial expression pictures, which can be said has a small number, were successfully trained in CNN and achieved high classification accuracy. Using the Haar Cascade library, the effect of unimportant pixels which is outside facial expressions were reduced. As different problems would require different network architectures it is required to figure out which architecture is the best for a particular problem. Though the proposed architecture has achieved a commendable result, it needs some improvements in some areas like adding more data in each class to get the more accurate result as it is known that deep learning is a data-driven approach. Emotion estimation from facial expressions is the area of interest of many researchers. It is hoped that this study will be a source of studies that will help in the early detection of diseases from facial expressions and also studies of consumer behavior analysis.

REFERENCES

[1] J. Zhao, X. Mao, and L. Chen (2019), "Speech emotion recognition using deep 1D \& 2D CNN LSTM networks," Biomedical Signal Processing and Control, vol. 47, pp. 312-323.
[2] M. Cabanac, "What is emotion?," Behavioural processes, vol. 60, pp. 69-83, 2002.
[3] T. Ahmed, S. Hossain, M. Hossain, R. Islam, and K. Andersson (2019), "Facial Expression Recognition using Convolutional Neural Network with Data Augmentation," pp. 1-17.
[4] D. Lundqvist, A. Flykt, and A. Öhman (1998), "The Karolinska directed emotional faces (KDEF)," CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, vol. 91, p. 630.
[5] K.-Y. Huang, C.-H. Wu, Q.-B. Hong, M.-H. Su, and Y.-H. Chen (2019), "Speech Emotion Recognition Using Deep Neural Network Considering Verbal and Nonverbal Speech Sounds," in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5866-5870: IEEE (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
[6] N. Christou and N. Kanojiya (2019), "Human facial expression recognition with convolution neural networks," in Third International Congress on Information and Communication Technology. Springer, pp. 539–545.
[7] X. Glorot, A. Bordes, and Y. Bengio (2011), "Deep sparse rectifier neural networks," in Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp. 315–323..
[8] M. Pantic and L. J. Rothkrantz (2000), "Automatic analysis of facial expressions: The state of the art," IEEE Transactions on Pattern Analysis \& Machine Intelligence, no. 12, pp. 1424–1445.
[9] M. G. Calvo and D. Lundqvist (2008), "Facial expressions of emotion (kdef): Identification under different display-duration conditions," Behavior research methods, vol. 40, no. 1, pp. 109–115.
[10] M. S. Hossain, S. Rahaman, R. Mustafa, and K. Andersson (2018), "A belief rule-based expert system to assess suspicion of acute coronary syndrome (acs) under uncertainty," Soft Computing, vol. 22, no. 22, pp. 7571–7586.
[11] Ahmad R. Naghsh-Nilchi and Mohammad Roshanzamir "An Efficient Algorithm for Motion Detection Based Facial Expression Recognition using Optical Flow" International Scholarly and Scientific Research and Innovation, 2008.
[12] T. Chang, G. Wen, Y. Hu, and J. Ma, "Facial Expression Recognition Based on Complexity Perception Classification Algorithm," arXiv e-prints, February 01, 2018